This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2929740, IEEE Transactions on Industrial Informatics

Deep Learning in Edge of Vehicles: Exploring Tri-Relationship for Data Transmission

Zhaolong Ning, Yufan Feng, Mario Collotta, Xiangjie Kong, Xiaojie Wang, Lei Guo, Xiping Hu, Bin Hu

***Abstract*—Currently, vehicles have the abilities to communicate with each other autonomously. For Internet of Vehicles (IoV), it is urgent to reduce the latency and improve the throughput for data transmission among vehicles. This paper proposes a deep learning based transmission strategy by exploring tri-relationships among vehicles. Specifically, we consider both the social and physical attributes of vehicles at the edge of IoV, i.e., edge of vehicles. The social features of vehicles are extracted to establish the network model by constructing triangle motif structures to obtain primary neighbors with close relationships. Additionally, the connection probabilities of nodes based on the characteristics of vehicles and devices can be estimated, by which a content sharing partner discovery algorithm is proposed based on Convolutional Neural Network (CNN). Finally, the experiment results demonstrate the efficiency of our method with respect to various aspects, such as message delivery ratio, average latency, and percentage of connected devices.**

***Index Terms*—Edge of vehicles, deep learning, triangle motif, device-to-device, data transmission.**

I. Introduction

With the rapid development of Internet of Things (IoT), there will be nearly 50 billion devices interconnected by 2020 [1]. Due to the advantage of short-distance wireless communication technologies and the widely usage of mobile devices, seamless interconnections among devices are gradually established [2], [3]. Internet of Vehicles (IoV) appears as a typical case of industrial IoT, in which ubiquitous information and messages can be exchanged and shared among vehicles without human interventions. This provides vehicle suppliers and service providers with unprecedented opportunities and enables them to develop emerging applications with rich multimedia contents, such as online games and traffic monitoring [4], [5]. However, new challenges also rise in data transmission and sharing in IoVs. Specifically, the rapidly growing demands of data transmission rates are restricted by the limited network bandwidth in the current cellular networks [6].

Device-to-Device (D2D) communication can play an intrinsic part in IoVs, since mobile devices are the main carriers for various mobile services [7]. It is regarded as an essential technology in 5G systems, which is promising to largely improve the spectrum efficiency and enable large-scale live video streaming by short-distance communications [8]. Devices can not only communicate with others autonomously following multihop manners without a centralized control, but also allow direct data transmission over proximate peer-to-peer links with the aid of centralized infrastructures [9]. D2D communication has been demonstrated to increase network capacities by spectrum reuse [10]. The ability of collecting real-time information is crucial to improve the efficiency of data transmission in IoVs [11], [20].

Generally speaking, D2D communication allows direct data transmission among vehicles and devices with similar directions or destinations. Thanks to the benefits achieved by the hop gains, proximity gains and spectrum reusing gains, data transmission latency can be largely reduced and data gathering efficiency can be significantly improved [12]. More than that, the Vehicle-to-Infrastructure (V2I) based data offloading can also be fulfilled effectively via D2D based links. It is noticed that the locations of mobile devices are consistent with the human movement trajectories. For instance, vehicles toward the same direction generally require similar contents, such as real-time traffic and road information. Vehicles equipped with wireless communication capabilities are promising to realize intelligent D2D communications. For example, a cluster-based resource block sharing and power allocation scheme is proposed in [13] based on D2D communications in IoVs.

Since the connections and information gathering of vehicles depend on the behaviors of human beings and the communication abilities of mobile devices, some researches explore social attributes in IoVs [14]. A vehicular communication framework is established in [15] by considering various social characters. A social-aware friend recommendation system is constructed in [16], where social behaviors are established based on the keywords of user interests in IoVs. In [17], a cooperative delay-tolerant content transmission strategy is introduced, aiming to share contents and minimize the cellular traffic loading. More than that, some social characteristics in IoVs, such as social communities, clustering coefficients and centralities, have been demonstrated to improve the efficiency of D2D communications [10], illustrating the relationships and internal social patterns among vehicles. In [18], authors discuss the recent radio resource management schemes to investigate the potential efficiency of D2D communications in safety-critical vehicle-to-everything (V2X) networks. Although D2D based communications have been widely discussed in IoVs, deep learning enabled solutions are not fully discussed, which can be viewed as a promising technique to manage network resources in dynamic scenarios [19].

Z. Ning, X. Hu and B. Hu are with the School of Information Science and Engineering, Lanzhou University, 730000, Lanzhou, China.

Z. Ning, Y. Feng, X. Kong (Corresponding author) and X. Wang are with the School of Software, Dalian University of Technology, 116620, Dalian, China. Email: xjkong@ieee.org.

Z. Ning and L. Guo are with the Chongqing Key Laboratory of Mobile Communications Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China.

Z. Ning is also with the State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China.

M. Collotta is with the Kore University of Enna, Enna 94100, Italy.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2929740, IEEE Transactions on Industrial Informatics

By jointly considering the social and physical characteristics, this paper proposes a deep learning based data transmission scheme by exploring tri-relationships among vehicles at the edge of networks (i.e., edge of vehicles), with the objective of improving the efficiency of D2D based communications. First, device connections are established according to their social features in the social layer. Second, devices are clustered by triangle motifs with a low-time complexity. After that, the connection probabilities are calculated, by which devices can obtain the data sharing partners through a deep learning based peer discovery scheme.

The main contributions of this paper can be summarized as follows:

- We consider both the social and physical characteristics of devices in the edge of vehicles. A connection framework is established to evaluate the interactions among vehicles by considering the tri-relationships from trajectory features.
- We design a triangle motif based clustering algorithm to control the network size. It can improve the efficiency of message transmission by eliminating the influences caused by accidental connected devices.
- We present a deep learning based neighbor selection method according to the probability of device connection, which can be calculated based on the mobilities of vehicles.
- We carry on the experiments on a real data set based on the bus trajectories of Hangzhou city (China), in October 2018. The results demonstrate the effectiveness of our method.

The rest of this paper is organized as follows. Section II illustrates the system model. The deep learning based content transmission strategy is specified in Section III. Performance evaluation is conducted in Section IV. Finally, we conclude this paper in Section V.

II. System Model

In this section, we illustrate the D2D based network model in Fig. 1, containing the social and physical layers in the edge of vehicles. Devices can be either content providers (d_i^p) or requesters (d_j^r). We focus on matching the tuples of d_i^p and d_j^r to improve the transmission efficiency. Specifically, once a transmission tuple is formed, the connection requirements from both social and physical layers should also be satisfied. The main notations are illustrated in Table I.

A. Physical Layer Model

The considered network includes m devices, which can be either content requesters or providers, denoted by $D^r = \{d_1^r, d_2^r, \cdots, d_m^r\}$ and $D^p = \{d_1^p, d_2^p, \cdots, d_m^p\}$, respectively. The uplink spectrum sharing model is considered in content transmission among devices. Generally, each D2D communication tuple can only reuse one uplink. In that case, how to obtain the real-time state of each transmission channel is challenging [21]. The establishment of D2D communication links depends on the connection probabilities of mobile nodes in edge of vehicles. It means as long as a physical connection

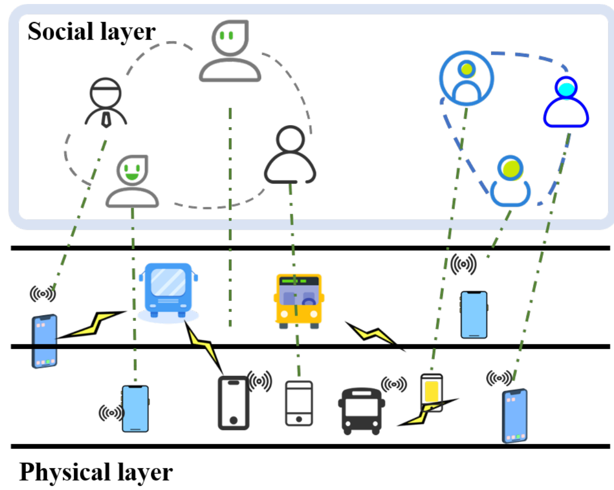


Fig. 1. The architecture of edge of vehicles with social and physical layers.

TABLE I
Main Notations

Notation	Definition
D	Device set
d_{ij}	Content transmission link between devices d_i and d_j
V	Vehicle set
R_{COM}	The maximal range of D2D communications
L	Routine set of vehilces
w_{ij}	Connection probability of d_i and d_j in the social layer
λ	Connection threshold in the social layer
f_{ij}	Connection probability of d_i and d_j in the physical layer
γ	Connection threshold in the physical layer
$En_{d_i l}$	Correlative vector of d_i and vehicle routines
$En_{d_i v}$	Correlative vector of d_i and vehicles

link d_{ij} between mobile devices d_i and d_j is established, the corresponding physical connection probability f_{ij} should be larger than threshold γ.

B. Social Layer Model

Vehicle users toward the same direction or destination tend to request similar contents. Generally, the content preferences of users change slowly, while the real-time traffic changes rapidly. Therefore, the social connection information can be collected and analyzed in an offline manner [22]. In this paper, the device connection in edge of vehicles is established based on the correlationship among vehicles and vehicle driving routines. Then, based on the device connections, the completed triangle motif is defined to extract the relationships among mobile devices in an efficient way.

1) Tri-Relationship Embedding: Various entities exist in the edge of vehicles, including vehicles, human beings, mobile devices, infrastructures and various fixed routines. The relationships among entities tend to be complex. Furthermore, there also exist tanglesome relationships among entities from different layers. Generally, the locations of mobile devices are consistent with the trajectories of human beings. Therefore,

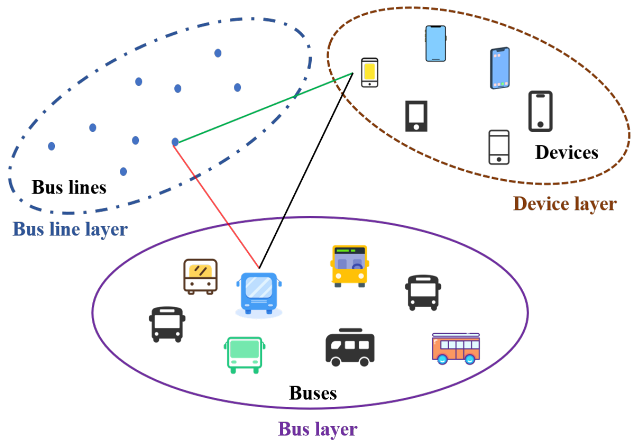


Fig. 2. The architecture of heterogenous entities in edge of vehicles.

the locations of human beings are utilized to represent the counterpart of time-varying devices.

This paper considers the mobile devices in buses. For simplicity, each passenger in the bus is assumed to carry an intelligent device. The device set can be denoted by $D = \{d_1, d_2, \dots, d_m\}$; the bus set is $V = \{v_1, v_2, \dots, v_n\}$; and the fixed bus line set is $L = \{l_1, l_2, \dots, l_k\}$. In general, one passenger with a device can only be in one bus, and one bus can only travel on one fixed bus line at a certain moment. According to the real bus data set, there are more than one bus running in each bus line.

The heterogeneous network is illustrated in Fig. 2, including the bus line, device and bus layers. Tri-relationship combines the location characteristics of vehicles and devices in the physical layer (as illustrated in Fig. 1) to construct the device-based subnetwork in Fig. 2. We define bus line engagement en_{d_il} to illustrate the duration that device d_i involves in one fixed bus line l. Similarly, the bus engagement can be expressed as en_{d_iv}. Variables en_{d_il} and en_{div} can be obtained according to the travelling behaviors of passengers. Therefore, the investigated network can be modelled as heterogeneous network $G = \{D, V, L\}$, in which D, V, L represent the node sets of devices, buses and fixed bus lines, respectively. Network social layer can be established by device engagement e_{dl} and en_{dv}. For mobile device d_i, it maintains two correlative vectors, i.e., $En_{d_il} = \{en_{i1}, en_{i2}, \dots, en_{ik}\}$ and $En_{d_iv} = \{en_{i1}, en_{i2}, \dots, en_{in}\}$, representing its bus line engagement and bus engagement, respectively. Based on the above vectors, the social connection can be obtained. Equation (6) presents the edge weight w_{ij} of devices d_i and d_j.

$$w_{ij} = \frac{En_{d_il} \times En_{d_jL}}{2 \times |En_{d_il}||En_{d_jl}|} + \frac{En_{d_iv} \times En_{d_jv}}{2 \times |En_{iv}||En_{jv}|}. \quad (1)$$

If $w_{ij} > \lambda$ holds, an edge between d_i and d_j can be constructed, where λ is a defined threshold and λ together with w_{ij} is between 0 and 1. Then, a weighted device connection subnetwork $GD = \{D, E_{GD}\}$ can be constructed by Equation (6), where D and E_{GD} represent the device set and the edge set of GD, respectively.

2) Device based Triangle Motif Definition: In order to accelerate the device connection estimation in the social layer, the minimal elementary units are defined as triangle motifs, representing the connections among three device entities. We adopt the definition of completed triangle motif in [23], and explain its definition briefly.

The completed triangle motif can be defined by tuple $\{M_{Tri}, A\}$, in which M_{Tri} is a binary adjacent matrix and A is an anchor node sequence. A simple example of device-based subnetwork is illustrated in Fig. 3. Here, Tri is a completed triangle motif containing three devices, and its corresponding anchor node set A and adjacent matrix M_{Tri} are constructed in an order of $\{d_p, d_q, d_r\}$, where $d_p \neq d_q \neq d_r$, and $d_p, d_q, d_r \in D$. Fig. 3 depicts a seven-node subnetwork GD, whose completed triangle motif-based representation is denoted by GD_{Tri}. It can be regarded as the enumeration of all the triangle structures satisfying M_{Tri} and A. Hence, we have $GD_{Tri}\{M_{Tri}, A\} = \{(d_1, d_5, d_3), (d_1, d_6, d_4), (d_2, d_5, d_4), (d_2, d_7, d_3)\}$.

C. Problem Formulation

In order to achieve high-efficient content transmission, all features from both social and physical layers are utilized to characterize the influences on device connection probabilities and social relationship closeness on content transmission. Our objective is to minimize the transmission delay for edge of vehicles. We solve this problem from both social and physical perspectives. In the social layer, the target content sharing devices are preferred to work in a D2D manner rather than broadcasting. Besides, content transmission for channel occupation should be taken into consideration. Content sharing links can be established as long as the results of connection estimation and social relationship are both above the corresponding threshold. Here, we define an indicator function $\xi(x|x_0)$ in Equation (2), where x is a variable:

$$\xi(x|x_0) = \begin{cases} 0 & x < x_0, \\ 1 & \text{otherwise.} \end{cases} \quad (2)$$

The device and vehicle sets are $D = \{d_1, d_2, \cdots, d_m\}$ and $V = \{v_1, v_2, \cdots, v_n\}$, respectively. We define a weighted matrix $W = \{w_{ij}\}_{m \times m}$, where w_{ij} is the weight of the edge between d_i and d_j. Hence, a D2D connection d_{ij} can be formed, as long as $w_{ij}\xi(w_{ij}|\lambda)\xi(f_{ij}|\gamma)$ is larger than 0.

For one device d_i, its transmission request can be satisfied only if two procedures are both finished, i.e., content retrieving and transmission. Since d_i retrieves requesting content from its neighbors with different probabilities, it results in different time cost. We define the content retrieving latency as t^0_{ij} for link d_{ij} by content sharing. The corresponding transmission latency of d_{ij} is t_{ij}. The total transmission content of link d_{ij} is denoted by Q_{ij}. In order to minimize the overall latency,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2929740, IEEE Transactions on Industrial Informatics

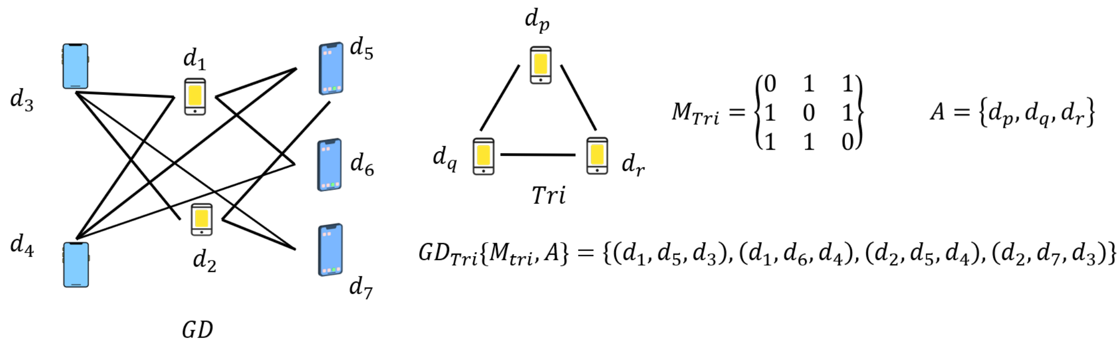


Fig. 3. An example of triangle motif definition: motifs Tri are leveraged to be detected in a simple device network GD on the left part. The motif is defined by a binary matrix M_{Tri} and an anchor set of nodes A.

the objective function is defined as follows:

$$\begin{aligned}
&\min \sum_{d_i \in D} \sum_{d_j \in D, d_i \neq d_j} (t_{ij}^0 + t_{ij}) m_{ij} \xi(w_{ij}|\lambda) \xi(f_{ij}|\gamma), \\
&s.t. \\
&\quad C_1 : \frac{Q_{ij}}{t_{ij}} \leq c, \\
&\quad C_2 : \sum_{d_i \in D} \sum_{\xi(w_{ij}|\lambda)\xi(f_{ij}|\gamma)>0}^{d_j \in D, d_i \neq d_j} Q_{ij} \geq Q_i^r, \\
&\quad C_3 : t_0, t_{ij} > 0, \\
&\quad C_4 : m_{ij} \in \{0, 1\}, \\
&\quad C_5 : \lambda, \gamma \in [0, 1], \\
&\quad C_6 : w_{ij}, f_{ij} \in [0, 1].
\end{aligned} \tag{3}$$

Constraint C_1 restricts the transmission rate cannot exceed channel capacity c. Constraint C_2 ensures that the received contents should not be less than the requested content Q_i^r. The remainders are some numerical constraints of variables mentioned above. Here, the subjective function aims to minimize the transmission latency by content sharing among links, i.e., discovering the neighbors containing the requesting content with high probabilities.

III. Deep Learning based Content Transmission in Edge of Vehicles

In this section, we specify the deep learning based content transmission strategy in edge of vehicles. It contains the connection probability estimation in the physical layer and the triangle motif based clustering in the social layer.

A. Triangle Motif based Clustering

The social layer intends to acquire the social tightness through exploring the social location information from a real data set. In general, the content preferences of users change slowly comparing with channel variations. Therefore, the social behaviors can be analyzed in an offline manner. On account of the huge data transmission and high-complexity preferences of user contents, the traditional parametric learning methods with fixed parameters can not be applied. Consequently, this paper proposes a motif based method to cluster the devices with tightness.

Based on the definition of triangle motif Tri, the corresponding motif based matrices, i.e., motif weighted matrix W_{Tri}, motif diagonal degree matrix D_{Tri}, and motif laplacian matrix $\daleth_{Tri}$, are expressed as follows:

$$(W_{Tri})_{ij} = \sum_{\{i,j,k\} \in GD\{M_{Tri}, A\}} w_{ij}, \ \{i, j\} \subset A, i \neq j, k \in V_{GD}, \tag{4}$$

$$(D_{Tri})_{ij} = \sum\nolimits_{j=1}^{n} (W_{Tri})_{ij}, \tag{5}$$

$$\daleth_{Tri} = I - D_{Tri}^{-1/2} W_{Tri} D_{Tri}^{1/2}. \tag{6}$$

For network g, it can be partitioned into two subnetworks, i.e., S and $\bar{S}$, where $g = S \cup \bar{S}$. The conductance between S and $\bar{S}$ in spectral clustering based on triangle motif is defined by:

$$\phi_{Tri}^{(g)}(S) = (cut_{Tri}^{(g)}(S, \bar{S}))/min(vol_{Tri}^{(g)}(S), vol_{Tri}^{(g)}(\bar{S})), \tag{7}$$

where $cut_{Tri}^{(g)}(S, \bar{S})$ is the number of triangle instances between S and $\bar{S}$. The value of $vol_{Tri}^{(g)}(S)$ illustrates the number of triangle instances, of which nodes belong to S. Then, we have $vol_{Tri}^{(g)}(S) = |S_{Tri}\{MTri, A\}|$. Similarly, $vol_{Tri}^{(g)}(\bar{S})$ can be calculated by $vol_{Tri}^{(g)}(\bar{S}) = |\bar{S}_{Tri}\{MTri, A\}|$.

Algorithm 1 is applied in the social layer to cluster the devices with high efficiency. The input of this algorithm is the subgraph set of the device connection subnetwork GD. The core idea is to select a cluster set of the input network by leveraging target motifs. According to the minimum conductance, the subgraphs in GD are clustered to obtain the closest ones, and a partition of nodes in S and $\bar{S}$ can be output. For a subgraph g, matrices W_{Tri}^g, D_{Tri}^g and $\daleth_{Tri}^g$ are calculated based on the steps from lines 3 to 6. After that, we can get the eigenvector with the second smallest eigenvalue of $\daleth_{Tri}^g$ in line 7. In line 8, the value of elements in $D_{Tri}^{g(-1/2)} z$ is reordered. Then, the triangle motif conductance becomes symmetric and satisfies $\phi_{Tri}^{(g)}(S) = \phi_{Tri}^{(g)}(\bar{S})$, so that any set of nodes (S and $\bar{S}$) can be interpreted as a cluster. As line 10 describes, the network is partitioned at edges, then the minimal conductance can be obtained. Two smaller subgraphs can be clustered and saved in the result set R.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2929740, IEEE Transactions on Industrial Informatics

Algorithm 1 The Triangle Motif based Network Cluster Algorithm

Input: Device network GD and triangle motif Tri
Output: Cluster set R of GD

/* R is initialized as an empty subgraph set.
for g in GD **do**
  $W^g_{Tri} =$ Number of instances of Tri
  $G^g_{Tri} =$ Weighted graph induced W^g_{Tri}
  $D^g_{Tri} =$ Diagonal matrix with $(D^g_{Tri})_{ii} = \sum_{j=1}^{n}(W^g_{Tri})_{ij}$
  $\urcorner_{Tri} = I - D^{g(-1/2)}_{Tri} W_{Tri} D^{g(-1/2)}_{Tri}$
  $z =$ Eigenvector of second smallest one for $\urcorner_{Tri}$
  $\sigma_i =$ Index of $D^{g(-1/2)}_{Tri} z$ with ith smallest value
  /* Sweep over all prefixes of σ
  $S = argmin_l \ \phi^{(g)}_{Tri}(S_l), \ where \ S_l = \sigma_1, \cdots, \sigma_k$
  Add S and $\bar{S}$ to R
end for
return R

With the above steps, the close devices can be gathered in one cluster, where devices inside have higher priorities for content transmission.

B. Connection Probability Estimation

The connection probabilities of devices have been investigated based on the mobilities of vehicles [16], [24]. In the social layer, connected edges in clusters can be obtained in Section II-B. In the physical layer, we define the communication range of D2D as R_{com}. If two devices are on one bus, they can share content information directly. Connection time T includes the on-bus time T_0 and the time out of bus within the communication coverage R_{com}, satisfying $T = T_0 + \{\min t | -R_{com} < Dis(x) < R_{com}\}$. Herein, variable t is the longest duration time of d_{ij}, and $Dis(x)$ is the real-time distance between two devices at time x, where $0 \le x \le t$ holds. If one passenger gets off the bus, the distances of devices depend on bus velocities. Otherwise, if more than one passenger gets off the bus, the distances among these devices are decided by the speeds of passenger themselves. For connection e_{ij} (devices d_i and d_j in the device-based subnetwork), the means and variances of their velocities are v_i and σ_i^2 for d_i, and v_j and σ_j^2 for d_j, respectively. Variable $Dis(x)$ is modelled as a Wiener process and the corresponding drift and variance are $\mu = v_i - v_j$ and $\sigma^2 = \sigma_i^2 + \sigma_j^2$, respectively. Hence, in infinitesimal time interval Δt, the increment of $Dis(t)$ follows a normal distribution, expressed by:

$$\Delta Dis(t) = Dis(T + \Delta t) - Dis(T) = \mu \Delta t + \sigma H, \tag{8}$$

where H corresponds to the normal distribution with mean value zero and variance Δt.

Since the Probability Density Function (PDF) of time evolution of particle velocity in Winer process can be described through Kolmogorov equation, we can obtain:

$$\frac{\partial p(r|dis_0, t)}{\partial t} = -\mu \frac{\partial p(r|dis_0, t)}{\partial r} + \frac{1}{2}\sigma^2 \frac{\partial^2}{\partial^2 r} p(r|dis_0, t), \tag{9}$$

where $r \in (-R_{com}, R_{com})$, and $p(r|Dis_0, t)$ is the PDF of $Dis(t)$ with $dis_0 = Dis(0)$. Let $\delta(\cdot)$ be the Dirac delta function, and it has the following initial and boundary conditions:

$$\begin{aligned} p(r|dis_0, 0) &= \delta(\cdot), \\ p(-R_{com}|dis_0, t) &= p(R_{com}|dis_0, t) = 0, t > 0. \end{aligned} \tag{10}$$

According to Equations (9) and (10), $p(r|dis_0, t)$ can be calculated by:

$$\begin{aligned} &p(r|dis_0, t) = \\ &\frac{1}{\sqrt{2\pi\sigma^2 t}} \sum_{y=-\infty}^{\infty} [\exp\{\frac{4y\mu R_{com}}{\sigma^2} - \frac{[(r - dis_0) - 4yL - \mu t]^2}{2\sigma^2 t}\} - \\ &\exp\{\frac{2\mu R_{com}(1-2y)}{\sigma^2} - \frac{[(r - dis_0) - 2R_{com}(1-2y) - \mu t]^2}{2\sigma^2 t}\}]. \end{aligned} \tag{11}$$

Therefore, the Cumulative Distribution Function (CDF) of total connecting T is:

$$F^{CDF}_{ij}(t) = Pr\{T \le t\} = 1 - \int_{-R_{COM}}^{R_{COM}} p(r|dis_0, 0) dr, \tag{12}$$

where variable $F^{CDF}_{ij}(t)$ is defined as the connection probability of d_i and d_j within time t. Then, we have $f_{ij} = F^{CDF}_{ij}(t)$.

C. Deep Learning based Transmission

In order to ensure the efficiency of content transmission, the requested content has no need to be broadcasted if device d_i contains the content requested by d_j. In order to achieve the above purpose, edges should be established based on the connection probability in each cluster. We leverage a Convolutional Neural Network (CNN) based method for probability estimation, which can automatically learn the filters based on the current training set. Comparing with other methods, CNN can generate outputs fastly and guarantee the result accuracy. The CNN-based framework is illustrated in Fig. 4. The training data set with hyper features is imported to the input layer of CNN, on which filters are utilized. Specifically, the hyper features in the input data set present the travelling behaviors, including the characteristics of real-time and historic travelling. In this paper, features include the bus engagement, bus line engagement, the corresponding travelling time, velocities and directions, describing the dynamic IoVs in the edge of vehicles. In addition, the input features can be changed varying different data sets and scenarios.

In the designed CNN model, the filter slides over the output to generate an intermediate hidden layer, where a set of non-overlapping partitions over the input are generated by the pooling operation. After that, a fully connected layer can be constructed to obtain the output according to the results of each hidden layer neuron.

Three parameters, i.e., the filter size, the stride size and the output size, are utilized in the CNN-based framework. Parameter filter size is decided by the value of n in the n-gram model. Considering the tri-relationships and completed triangle motif in the edge of vehicles, we define $n = 3$ to refine the device clusters, indicating the number for one device with its bus and bus line engagements. The stride size

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2929740, IEEE Transactions on Industrial Informatics

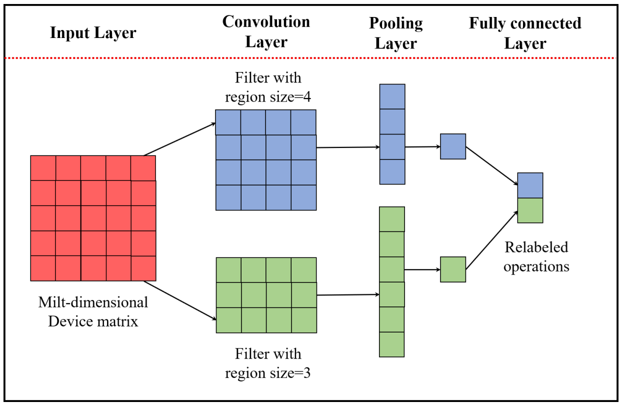


Fig. 4. The CNN framework used to establish the probabilities among devices.

determines the displacement of each step. The smaller the step size is, the more overlapping area of the continuous filters is. Therefore, high-precision results with less overlapping area can be obtained by the large step size. Output size τ_{out} is relative to the size of the input layer. In order to achieve a high accuracy value, τ_{out} can be expressed as:

$$\tau_{out} = (\tau_{in} + 2\tau_{padding} - \tau_{filter}) + 1. \tag{13}$$

Herein, variables τ_{in}, $\tau_{padding}$ and τ_{filter} are the input size, the padding size and the filter size applied in the input layer of the CNN model. The wide convolution (zero-padding) is embedded, and the padding size is defined as $\tau_{padding} = (\tau_{filter} - 1)/2$. This guarantees that filters can be applied to each element of the input network. After that, filters can produce larger outputs than those of the input layer. Pooling is a feature extraction layer commonly used after convolution in CNN, to integrate features pointed in small neighbors to get new features. On one hand, it prevents useless parameters from increasing the time complexity; on the other hand, it increases the integration of features. Algorithm 2 specifies the above procedures integrating with the connection probability estimation.

Algorithm 2 Deep Learning based Transmission Model

Input: Clustered devices with hyper features

```
Import the data set and divide it into the training set and the testing set
Initialize τ_filter, τ_padding and τ_stride
Define CNN(τ_filter, τ_padding, τ_stride, sigmoid(x))
Load bus-bus line area labels
/* Relabel devices with bus-bus line labels
P{label_q|device} = P{label_q}P{device|label_q}
if P{label_q|d_i} > γ then
    for d_j that connects d_i do
        P_ij = max {P_i + P_j, P_ij}
    end for
    Remain edge e_ij, and obtain connection probability f_ij.
end if
Handle content requests of devices based on f_ij.
```

This probability estimation algorithm takes the device clusters from the social layer as input and relabels the devices with CNN, aiming to refine the size of each cluster. Then, devices can transmit messages to their neighbors according to connection probabilities.

D. Complexity Analysis

We analyze the computational complexity of our method in this subsection. The main time-consuming procedures are device based subnetwork establishment and triangle motif based clustering. The deep learning based procedure can generate the real-time output in an efficient way. Moreover, the inputs of deep learning procedure are in small sizes. Therefore, we do not discuss this procedure here.

Device connection subnetwork establishment procedure: Considering there exists m devices, n buses, and k bus lines, the weights of device based subnetwork can be obtained by $O(\max\{n^2, k^2\})$. The number of buses is generally larger than that of the bus line in one city. Hence, the complexity of weight calculation is $O(n^2)$. For m devices, it can be worked out by $O(mn^2)$.

Triangle motif based clustering procedure: Generally, the time complexity of the triangle motif based algorithm depends on the construction of the motif-based adjacency matrix and the corresponding eigenvectors. Here, we can traverse the subnetwork GD with the complexity $O(m^2)$. For the corresponding device based matrices, we can construct the corresponding motif based representation GD_{Tri}. For the completed triangle motif with three fully connected nodes, the motif based adjacent matrix W_{Tri} can be worked out within $O(m^3)$ in a complete graph with m nodes. Considering there are u edges in the device-based subnetwork, we can calculate the eigenvectors through the motif-based laplacian matrix by $O((u+n)(\log m)^{O(1)})$. After that, the complexity of sorting the eigenvector indexes becomes $O(m \log m)$. Therefore, the computation complexity of the triangle motif based clustering procedure is $O(m^3)$.

In summary, the time complexity of the whole approach proposed is $O(\max\{mn^2, m^3\})$. Generally, the number of devices is larger than that of vehicles. Hence, our method can be fulfilled within $O(m^3)$.

IV. Performance Evaluations

In this section, we evaluate our method based on Python and Matlab. Experiments are carried on the real data set of bus trajectories in October 2018, Hangzhou (China). It includes the information of the bus ID, bus line, bus driving time, and passenger bus card. The ID of passenger bus card is unique. The information of bus trajectories, passenger engagements and bus lines can be obtained from the data set. We first illustrate the experiment configurations. Then, the experiment results are demonstrated and analyzed.

A. Experiment Settings

The mean velocities of vehicles are assumed to be uniformly distributed within between 0 and 50 km/h. The threshold values of λ (in Section II-B) and γ (in Section III-B) are trained first. In this experiment, we evaluate our method from the following two aspects:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2929740, IEEE Transactions on Industrial Informatics

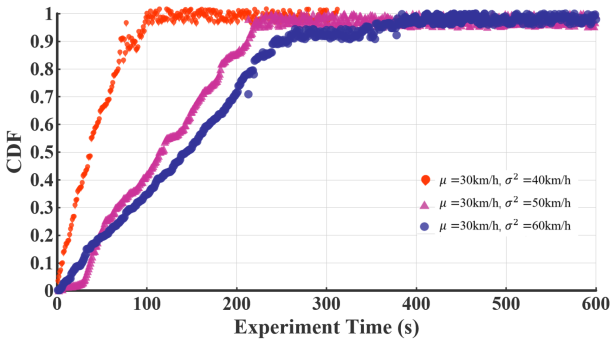


Fig. 5. The values of CDF versus different means μ and variances σ^2 of vehicles ($\lambda = 0.4$).

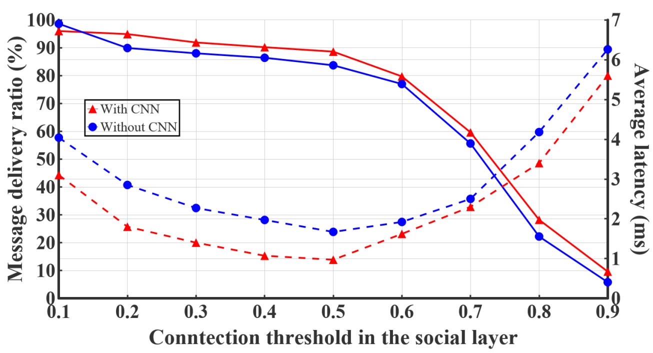


Fig. 6. The effect of connection threshold λ effects on message delivery ratio and average latency ($\gamma = 0.4$).

- Message delivery ratio: It is the ratio of successfully transmitted messages to the overall number of messages in the edge of vehicles.
- Average latency: It is the average duration time, which starts from the moment that the required message is created and ends when that message is delivered successfully.

To demonstrate the effectiveness of our method from the above two perspectives, we compare it with two schemes:

- Peer discovery scheme [25]: D2D users in the network are first grouped according to the social community and centralities. Then the optimal beacon is decided and sent by each group at constant intervals.
- Social-aware approach [26]: The physical-layer device network is first formulated by the encounter information. Then the content is transmitted based on their contents by links.

According to the description in Section II-B, connections formed in the social layer depend on threshold λ. A larger threshold λ results in less social device connections, smaller social device clusters and less content transmission links. Since the change of γ in the physical layer also reflects the performance of these two metrics, we first train the parameters of λ and γ.

B. Result Analysis

Fig. 5 illustrates the curve changes of $f_{ij}(t)$ towards the means and variances of the device velocities. As the value of

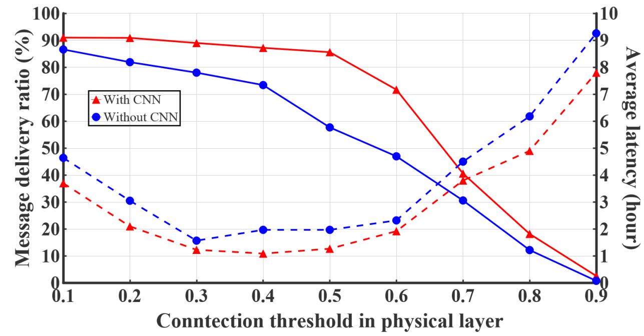


Fig. 7. The effect on threshold γ on message delivery ration and average latency in the physical layer ($\lambda = 0.4$).

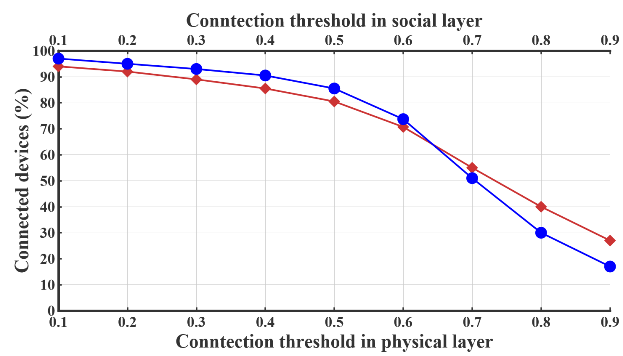


Fig. 8. The percentage of connected devices versus physical threshold γ ($\lambda = 0.4$) and social threshold λ ($\gamma = 0.45$).

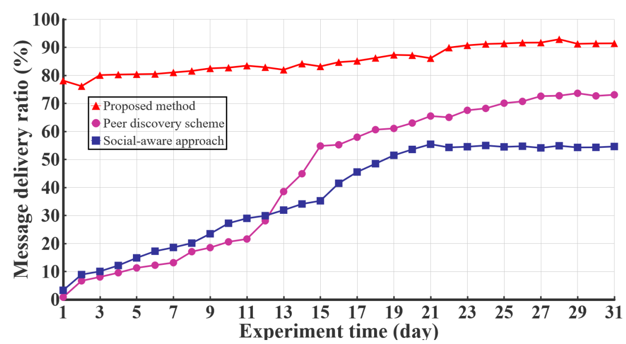


Fig. 9. Comparison of message delivery ratio versus experiment time ($\lambda = 0.4, \gamma = 0.45$).

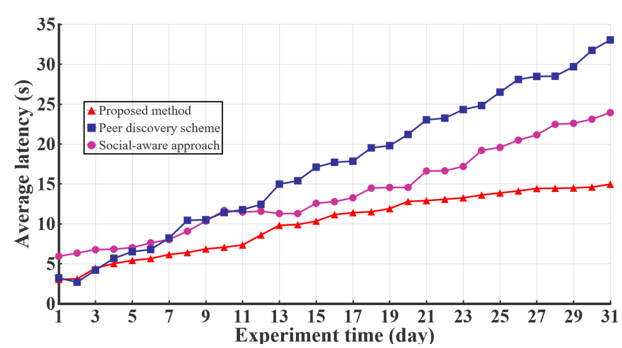


Fig. 10. Comparison of average latency versus experiment time ($\lambda = 0.4, \gamma = 0.45$).

σ^2 increases, the curve of $F_{ij}(t)$ expands in width. We can observe that the probabilities of device disconnection enhance as the gap between velocities increases. We select the first week in October, 2018 to train the two thresholds, i.e., λ and γ. First, we set $\gamma = 0.4$ and measure the effects of λ by considering both two metrics, i.e., message delivery ratio and average latency. Then, in order to evaluate the effectiveness of CNN-based procedure in our approach, we compare two cases, i.e., our proposed method with CNN and without CNN, when conducting parameter training processes.

In Figs. 6 and 7, the solid lines correspond to the y-axis of message delivery ratio, and the dashed lines reflect the average latency changes with the left y-axis. We can observe that when $\lambda \in [0.2, 0.6]$, the corresponding message delivery ratio decreases slowly. However, if the value of λ continues to increase (larger than 0.6), the message delivery ratio declines quickly. On the contrary, the average latency first decreases slowly and then increases at a gentle rate in the proposed method both with CNN and without CNN. Results demonstrate that our CNN-based method achieves better performances in both message delivery ratio and average latency.

Fig. 8 illustrates the percentage of connected devices changes with different values of γ and λ. First, we set $\lambda = 0.4$ and observe that the percentage of connected vehicles declines sharply when $\gamma > 0.5$ (blue line). When $\gamma \in [0.1, 0.5]$, the connected vehicles are in $[100\%, 80\%]$. Then, when we set $\gamma = 0.45$, it shows that message delivery ratio and average latency achieve better performances when γ is restricted within $[0.2, 0.5]$. By parameter training, the restrictions from social and physical layers are considered to remove the accident connections and decline the content retrieving delay from neighbors. However, the message delivery ratio decreases due to less number of device connections. By comprehensively considering the obtained results, we define $\lambda = 0.4$ and $\gamma = 0.45$ in the following experiments. It can make a balance of the two metrics, so that a low latency can be achieved with little message delivery loss.

We further compare our method with two algorithms, i,e., the peer-discovery scheme and the social-aware approach. In this part, we analyze the whole traffic data set in October, 2018 to verify the effectiveness of our method. As Fig. 9 shows, the message delivery ratios of three algorithms change with the experiment time. The results elaborate that our proposed method achieves the best performance. The message delivery ratio increases due to the accumulative bus and bus line learning process. The average latency results are demonstrated in Fig. 10. We can observe that the average latency increases with the experiment time, in terms of the increasing number of devices and the updated offline social information. Our method obtains the lowest average latency comparing with the other two approaches.

V. Conclusion

In order to improve the efficiency of content transmission, a deep learning based transmission scheme is put forwarded by exploring tri-relationship among nodes. We consider both the social and physical characteristics of D2D communications in edge of vehicles. The correlative trajectory features of devices and buses are extracted to establish the device-based subnetwork model, which can be clustered with triangle motif structures to obtain primary close neighbors. After comprehensively considering the physical characteristics of buses and devices, we obtain the connection probabilities of devices, by which a content sharing partner discovery algorithm is proposed based on CNN. Finally, the experiment results verify the effectiveness of our method with respect to various performance metrics. In our future work, we will improve the efficiency of content transmission under the premise of considering content permissions.

VI. Acknowledgments

The work is supported by the National Nature Science Foundation of China under Grant 61632014 and Grant 61802159, China Postdoctoral Science Foundation under Grant 2018T110210, Fundamental Research Funds for the Central Universities under Grant DUT19JC18 and Grant DUT18JC09, State Key Laboratory of Integrated Services Networks, Xidian University (ISN20-01), and Science and Technology Innovation Program of National Defense.

References

[1] Z. Ning, J. Huang, and X. Wang, “Vehicular fog computing: Enabling real-time traffic management for smart cities,” *IEEE Wireless Communications*, vol. 26, no. 1, pp. 87–93, 2019.

[2] R. Deng, S. He, C. Peng, and Y. Sun, “Towards balanced energy charging and transmission collision in wireless rechargeable sensor networks,” *Journal of Communications & Networks*, vol. 19, no. 4, pp. 341–350, 2017.

[3] S. Huang, B. Li, B. Guo, J. Zhang, P. Luo, D. Tan, and W. Gu, “Distributed protocol for removal of loop backs with asymmetric digraph using gmpls in p-cycle based optical networks,” *IEEE Transactions on Communications*, vol. 59, no. 2, pp. 541–551, 2010.

[4] X. Wang, Z. Ning, and L. Wang, “Offloading in Internet of vehicles: A fog-enabled real-time traffic management system,” *IEEE Transactions on Industrial Informatics*, vol. 14, no. 10, pp. 4568–4578, 2018.

[5] X. Li, S. Yin, X. Wang, Y. Zhou, Y. Zhao, S. Huang, and J. Zhang, “Content placement with maximum number of end-to-content paths in k-node (edge) content connected optical datacenter networks,” *Journal of Optical Communications and Networking*, vol. 9, no. 1, pp. 53–66, 2017.

[6] Z. Ning, X. Wang, F. Xia, and J. J. Rodrigues, “Joint computation offloading, power allocation, and channel assignment for 5G-enabled traffic management systems,” *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 3058–3067, 2019.

[7] Y. Rong, J. Ding, X. Huang, M. T. Zhou, S. Gjessing, and Z. Yan, “Optimal resource sharing in 5G-enabled vehicular networks: A matrix game approach,” *IEEE Transactions on Vehicular Technology*, vol. 65, no. 10, pp. 7844–7856, 2016.

[8] R. Deng, R. Lu, C. Lai, T. H. Luan, and H. Liang, “Optimal workload allocation in fog-cloud computing toward balanced delay and power consumption,” *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 1171–1181, 2016.

[9] Z. Zhou, C. Gao, X. Chen, Z. Yan, S. Mumtaz, and J. Rodriguez, “Social big data based content dissemination in Internet of vehicles,” *IEEE Transactions on Industrial Informatics*, vol. 14, no. 2, pp. 768–777, 2018.

[10] Z. Ning, F. Xia, X. Hu, Z. Chen, and M. S. Obaidat, “Social-oriented adaptive transmission in opportunistic internet of smartphones,” *IEEE Transactions on Industrial Informatics*, vol. 13, no. 2, pp. 810–820, 2017.

[11] F. Zhang, R. Deng, and L. Hao, “An optimal real-time distributed algorithm for utility maximization of mobile ad hoc cloud,” *IEEE Communications Letters*, vol. 22, no. 4, pp. 824–827, 2018.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2929740, IEEE Transactions on Industrial Informatics

[12] W. Sun, E. Strom, F. Brannstrom, K. Sou, and Y. Sui, “Radio resource management for D2D-based V2V communication,” *IEEE Transactions on Vehicular Technology*, vol. 65, no. 8, pp. 6636–6650, 2016.

[13] W. Sun, Y. Di, E. Strom, and F. Brannstrom, “Cluster-based radio resource management for D2D-supported safety-critical V2X communications,” *IEEE Transactions on Wireless Communications*, vol. 15, no. 4, pp. 2756–2769, 2016.

[14] Z. Ning, F. Xia, N. Ullah, X. Kong, and X. Hu, “Vehicular social networks: Enabling smart mobility,” *IEEE Communications Magazine*, vol. 55, no. 5, pp. 16–55, 2017.

[15] Z. Ning, X. Hu, Z. Chen, M. Zhou, B. Hu, J. Cheng, and M. S. Obaidat, “A cooperative quality-aware service access system for social Internet of vehicles,” *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2506–2517, 2018.

[16] T. H. Luan, X. Shen, B. Fan, and L. Sun, “Feel bored? join verse! engineering vehicular proximity social networks,” *IEEE Transactions on Vehicular Technology*, vol. 64, no. 3, pp. 1120–1131, 2015.

[17] G. Mao, Z. Zhang, and B. Anderson, “Cooperative content dissemination and offloading in heterogeneous mobile networks,” *IEEE Transactions on Vehicular Technology*, vol. 65, no. 8, pp. 6573–6587, 2016.

[18] L. Gallo and J. Haerri, “Unsupervised long- term evolution device-to-device: A case study for safety-critical V2X communications,” *IEEE Vehicular Technology Magazine*, vol. 12, no. 2, pp. 69–77, 2017.

[19] Z. Ning, P. Dong, X. Wang, J. Rodrigues, and F. Xia, “Deep reinforcement learning for vehicular edge computing: An intelligent offloading system,” *ACM Transactions on Intelligent Systems and Technology*, vol. 25, p. 1, 2019.

[20] X. W. Wang, Z. Ning, X. Hu, L. Wang, L. Guo, and B. Hu, “Future communications and energy management in internet of vehicles: Toward intelligent energy-harvesting, doi: 10.1109/mwc.2019.1900009,” *IEEE Wireless Communications*, 2019.

[21] H. Huang, S. Huang, S. Yin, M. Zhang, J. Zhang, and W. Gu, “Virtual network provisioning over space division multiplexed optical networks using few-mode fibers,” *IEEE/OSA Journal of Optical Communications and Networking*, vol. 8, no. 10, pp. 726–733, 2016.

[22] Z. Ning, L. Liu, F. Xia, B. Jedari, I. Lee, and W. Zhang, “Cais: A copy adjustable incentive scheme in community-based socially aware networking,” *IEEE Transactions on Vehicular Technology*, vol. 66, no. 4, pp. 3406–3419, 2016.

[23] A. R. Benson, D. F. Gleich, and J. Leskovec, “Higher-order organization of complex networks,” *Science*, vol. 353, no. 6295, p. 163, 2016.

[24] G. Yan and S. Olariu, “A probabilistic analysis of link duration in vehicular ad hoc networks,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, pp. 1227 – 1236, 2012.

[25] B. Zhang, L. Yong, D. Jin, H. Pan, and H. Zhu, “Social-aware peer discovery for D2D communications underlaying cellular networks,” *IEEE Transactions on Wireless Communications*, vol. 14, no. 5, pp. 2426–2439, 2015.

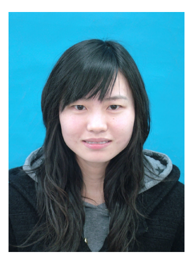
[26] Y. Zhang, L. Song, W. Saad, Z. Dawy, and Z. Han, “Exploring social ties for enhanced device-to-device communications in wireless networks,” in *IEEE GLOBECOM*, 2013, pp. 4597–4602.

Zhaolong Ning (M’14-SM’18) received the M.S. and Ph.D. degrees from Northeastern University, China, in 2011 and 2014, respectively. He is an Associate Professor with the Dalian University of Technology, and an Adjunct Professor with the Lanzhou University, China. He has published over 100 scientific papers in international journals and conferences. His research interests include mobile edge computing, vehicular network, and network optimization.

Yufan Feng received the BSc degree from Hebei University in 2016, and the M.S. degree from Dalian University of Technology, China, in 2019. She joined Baidu Inc after graduation. Her interest includes vehicular network and social network.

Mario Collotta is an Associate Professor of Computer Engineering - Faculty of Engineering and Architecture, Kore University of Enna, Italy. He was the Chair of the BD course in Computer Science Engineering and of the MD course in Telematics Engineering with the Kore University of Enna. He is scientific responsible of the Computer Engineering and Networks Laboratory. His research activity is mainly focused on the study of innovative solutions and approaches in expert systems and networks, focused on real-time and secure application.

Xiangjie Kong (M’13-SM’17) received the BSc and PhD degrees from Zhejiang University, Hangzhou, China. He is currently an Associate Professor in School of Software, Dalian University of Technology, China. He has served as (Guest) Editor of several international journals, Workshop Chair or PC Member of a number of conferences. Dr. Kong has published over 100 scientific articles in international journals and conferences (with 70+ indexed by ISI SCIE). His research interests include intelligent transportation systems, mobile computing, and cyber-physical systems.



Xiaojie Wang received the M.S. degree from Northeastern University, China, in 2011. From 2011 to 2015, she was a software engineer in NeuSoft Corporation, China. Currently, she received her Ph.D. degree from Dalian University of Technology, China, in 2019. Her research interests are vehicular networks, edge computing, resource management. She has published over 30 scientific papers in the above areas.

Lei Guo received the Ph.D. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 2006. He is currently a Full Professor with Chongqing University of Posts and Telecommunications, Chongqing, China. He has authored or coauthored more than 200 technical papers in international journals and conferences. He is an Editor for several international journals. His current research interests include communication networks, optical communications, and wireless communications.

Xiping Hu is currenlty a professor with Lanzhou University, China. He has over 70 papers published and presented in prestigious conferences and journals. He has been serving as the lead guest editors of IEEE Transactions on Automation Science and Engineering, WCMC and etc. His research areas consist of mobile cyber-physical systems, crowdsensing and social networks. He holds a PhD in The University of British Columbia, Vancouver, Canada.

Bin Hu (M'10–SM'15) is currently a professor in Lanzhou University, adjunct professor in Tsinghua University, China, and guest professor in ETH Zurich, Switzerland. He is also IET Fellow, co-chairs of IEEE SMC TC on Cognitive Computing, and Member at Large of ACM China, Vice President of International Society for Social Neuroscience (China committee) etc. His work has been funded as a PI by the Ministry of Science and Technology, National Science Foundation China, European Framework Programme 7, EPSRC, and HEFCE UK, etc, also, published more than 100 papers in peer reviewed journals, conferences, and book chapters including Science, Journal of Alzheimer's Disease, PLoS Computational Biology, IEEE Trans., IEEE Intelligent Systems, AAAI, BIBM, EMBS, CIKM, ACM SIGIR, etc. He has served as Chairs/Co-Chairs in many IEEE international conferences/workshops, and associate editors in peer reviewed journals on Cognitive Science and Pervasive Computing, such as IEEE Trans. Affective Computing, Brain Informatics, IET Communications, Cluster Computing, Wireless Communications and Mobile Computing, The Journal of Internet Technology, Wiley's Security and Communication Networks, etc.